Forecasting elections

Article

Accepted Version

Vaughan Williams, L. and Reade, J. J. ORCID: https://orcid.org/0000-0002-8610-530X (2016) Forecasting elections. Journal of Forecasting, 35 (4). pp. 308-328. ISSN 1099-131X doi: 10.1002/for.2377 Available at https://centaur.reading.ac.uk/47396/

Forecasting Elections

Leighton Vaughan Williams
Nottingham Business School[*]

J. James Reade
University of Reading[†]

April 21, 2015

Abstract

In this paper we assess opinion polls, prediction markets, expert opinion, and statistical modelling over a large number of US elections in order to determine which perform better in terms of forecasting outcomes. In line with existing literature, we bias-correct opinion polls. We consider accuracy, bias and precision over different time horizons before an election, and we conclude that prediction markets appear to provide the most precise forecasts and are similar in terms of bias to opinion polls. We find that our statistical model struggles to provide competitive forecasts whilst expert opinion appears to be of value. Finally we note that the forecast horizon matters; while prediction market forecasts tend to improve the nearer an election is, opinion polls appear to perform worse, while expert opinion performs consistently throughout. We thus contribute to the growing literature comparing election forecasts of polls and prediction markets.

JEL Classification: C53, D83, D72.
Keywords: Forecasting Models, Information and Knowledge, Elections, Voting Behaviour, Prediction Markets.

1 Introduction

There exist many sources of information one could use to forecast the outcome of an election *ex ante*; statistical models, expert opinion, opinion polls, and prediction markets are just four. Any such forecast is dependent on some set of information amassed at a particular point in time prior to the event happening, denoted $\mathcal{I}_t$, and also on the model through which that information is processed, $f_t(\mathcal{I}_t)$. In this paper we consider each of these potential models: *statistical models*, where information generally released by statistical agencies is processed using statistical methods, *expert opinion*, where information is processed by experts, *polls*, where information from potential voters is processed

[*]Nottingham Business School, Nottingham Trent University, Burton Street, Nottingham, United Kingdom, NG1 4BU, Email: leighton.vaughan-williams@ntu.ac.uk, Phone: +44 (0) 115 848 6150.

[†]Corresponding author. Department of Economics, HumSS Building, Whiteknights Campus, Reading, United Kingdom, RG6 6AA. Email: j.j.reade@reading.ac.uk, Phone: +44 (0) 118 378 5062.

by polling companies and released, and *prediction markets*, where agents may also use potentially private information to buy and sell contracts contingent on a particular future event, thus revealing information in the process of doing so. As such, we are providing an additional perspective on the so-called Hayek hypothesis (Hayek, 1945; Smith, 1982) which suggests that markets can work efficiently even when participants have a limited knowledge of the environment or other participants (see also Hurley and McDonough, 1995).

In doing so, we build upon prior literature which identifies different types of prediction market, classified according to type of contract (Snowberg and Zitzewitz, 2005), and which have sought to examine the historical accuracy of election markets (Rhode and Strumpf, 2004) and to compare and/or relate the behaviour and performance of these markets to that of opinion polls (e.g. Kou and Sobel, 2004; Leigh and Wolfers, 2006; Berg et al., 2008; Rothschild, 2009).

We make use of a vast, novel dataset from historical US elections to conduct our forecast comparison exercise. We assess the types of forecast based on the same criteria: Accuracy, bias, and precision of historical forecasts. By accuracy we mean how often a forecast correctly predicts the election outcome, by bias whether the expected vote share or outcome probability is equal to the actual vote share or true probability, and by precision the variance of forecast errors. We find that prediction markets appear to provide the most precise forecasts and are similar in terms of bias to opinion polls. We find that our statistical model struggles to provide competitive forecasts whilst expert opinion appears to be of value. Finally we note that the forecast horizon matters; while prediction market forecasts tend to improve the nearer an election is, opinion polls appear to perform worse, while expert opinion performs consistently throughout.

In Section 2 we introduce our object of interest, the outcome of an election before we introduce in Section 3 our candidate forecast models and the datasets we have for each forecast model. Section 4 then discusses the methodology we use in assessing these forecast models and Section 5 outlines our results. Section 6 concludes.

2 The Actual Outcome

The outcomes of an election are manifold; more often than not in US elections, there are two candidates (a Republican and a Democrat), and the vote share each receives is one outcome of interest, as well as who actually wins each election.[1]

We think of the two-party vote share for candidate or party i in election j as $V_{i,j,T}$, where T is the date of the election, and we denote forecasts of that vote share made by forecaster f at date t, where $t < T$, as $\widehat{V}_{i,j,f,T|t}$. If we are considering only the two-party vote share, then the alternative outcome of interest is whether or not $V_{i,j,T} > 0.5$, as in this situation party i has won the election in terms of vote share, and hence we might think of a binary variable:

$$W_{i,j,T} = 1_{\left\{V_{i,j,T}=\max_k\{V_{k,j,T}\}_{k=1}^{N}\right\}}. \tag{1}$$

[1] This is particularly so for US Presidential Elections which are determined by the electoral college system and hence anomalies like the 2000 Bush vs. Gore election can happen where the winner was Bush even though Gore gained the larger (popular) vote share.

That is, $W_{i,j,T}$ is 1 if party i wins election j (in terms of vote share), zero otherwise. We define $W_{i,j,T}$ in (1) generally for an N-candidate election, yet often elections in the US involve just two candidates, and in that situation the probability that $V_{i,j,T} > 0.5$, i.e. the vote share on election day is sufficient to win the election popular vote, is what matters. The forecast made at time $t < T$ of whether or not a vote share $V_{i,j,T}$ will be sufficient to win an election we denote as $\widehat{W}_{i,j,f,T|t} = \widehat{\mathsf{P}}_t \left(V_{i,j,T} = \max_k \{V_{k,j,T}\}_{k=1}^N \right)$, or $\widehat{W}_{i,j,f,T|t} = \widehat{\mathsf{P}}_t \left(V_{i,j,T} > 0.5 \right)$ in the case of a two-candidate election.

3 The Candidate Forecast Models

We consider a number of sources of pre-election forecasts in this paper:

1. Opinion polls as collated by *Real Clear Politics* and *Polling Report*, two websites that collect historical and current polling information surrounding elections.[2]
2. Price data from Iowa Electronic Markets, an online prediction market for various political (and other) events.[3]
3. Price data from Betfair, an online betting company that offers markets which include political events or else which include election outcomes.[4]
4. Price data from Intrade, an online betting company that offers predominantly political markets or perhaps politically related markets.[5]
5. Expert opinion as canvassed for the PollyVote Project.
6. Regression modelling using the methodology of Fair (1996) in order to forecast outcomes using macroeconomic variables.

We consider each to be a forecast model; a mechanism that transforms information available at time $t < T$, $\mathcal{I}_t$ into a forecast for either a vote share $\widehat{V}_{i,j,f,T|t}$ or a probability of the election outcome $\widehat{W}_{i,j,f,T|t}$. In the next four subsections we describe each of these sources of data and comment on the mechanisms that generate forecasts from information available at time t.

3.1 Opinion Polls

Opinion polls are conducted by numerous companies in the US surrounding all sorts of elections and political questions (e.g. presidential approval). In the case of elections, polls are forecasts of vote shares conducted at some point $t < T$ by polling company f, hence they are denoted as $\widehat{V}_{i,j,f,T|t}$. Notionally, polls reflect public opinion regarding voting for particular candidates, and assuming the sample upon which they are based is representative, they can be seen as some reflection of voting intentions at time t, something which we denote $V_{i,j,t}$. As such, to treat a poll as a forecast of the eventual election outcome, we assume

[2]See http://www.realclearpolitics.com/ and http://www.pollingreport.com/ for details.
[3]See http://tippie.uiowa.edu/iem/.
[4]See http://www.betfair.com.
[5]See http://www.intrade.com/.

thus that such voting intentions do not change in the intervening time period. Hence there are at least two sources of error: the first is that the vote share forecast by the poll ($\widehat{V}_{i,j,f,T|t}$) may not be a true reflection of $V_{i,j,t}$; and/or $V_{i,j,t}$ may differ substantially from $V_{i,j,T}$ due, for example, to the learning process that takes place during an election campaign on the part of voters.

Furthermore, political candidates are keen observers of polls and thus to some extent there may be endogeneity; candidates may respond to poll outcomes when $t < T$, increasing or decreasing effort levels. For example, a particularly disappointing set of polls may lead to a candidate increasing their effort in an election, which may thus impact $V_{i,j,T}$ causing it to differ from $V_{i,j,t}$. Furthermore, the success of campaign fundraising efforts may also be affected by poll outcomes (and potentially prediction markets also). As some polling companies are known or suspected to favour one political party or the other, there is the possibility of strategic behaviour on the part of pollsters in the timing and nature of their polls.

Nonetheless, considering the data at our disposal, we take polls to be forecasts of voting intentions *on election day*, T, as expressed at time t, and we analyse the extent to which they are effective forecasts of $V_{i,j,T}$. We take our data from *Real Clear Politics* (RCP), and *Polling Report*, websites which compile polling data from thousands of US elections over recent years.[6] Table 5 summarizes the elections from which we have collected data from RCP; overall we have 19,277 observations from 394 different elections ranging from presidential elections in 2004 and 2008 both at the national and state levels, senate, governor and house elections and also Republican presidential candidate selection processes in 2008 and 2012, and the Democratic selection process from 2008.

We collect information on the polling company, the length of time the poll was conducted over, size and type of audience polled (likely voters or registered voters), forecast vote share for each candidate, and we also record the final outcome of each election.[7] There are averages of polls that are constructed by various groups, such as RCP themselves, and also others such as Nate Silver at the blog *FiveThirtyEight*. Although averaging can be a useful tool, particularly if the weights are appropriate (see, e.g., Bates and Granger, 1969; Graefe et al., 2012), it can only outperform the best individual forecast within the pool of forecasts being averaged in the presence of systematic bias (for example if one forecast is known to be positively biased and another negatively biased). As we correct polls for systematic biases, the need for averaging within polls is mitigated. There remains a potential role for combining forecasts from our different sources, nonetheless, and we explore this further in Section **??**.

Figure 1 plots poll outcomes for Obama's vote share during 2008 for the 2008 Presidential election; his final 52.9% vote share is denoted by the solid black line. The plot should be viewed from right to left, as the horizontal axis is the number of days remaining until the election takes place. Different polling

[6]See http://www.realclearpolitics.com/ and http://www.pollingreport.com/ for our two sources.

[7]We have data on 446 individual poll producers, however many of the producers of polls are collaborations, such as Reuters and Zogby or Reuters and Ipsos. It is hard to get a precise number of the different forecasting companies involved because RCP often lists them abbreviated, but it appears there are around 200 distinct companies or organizations reflected in our dataset.

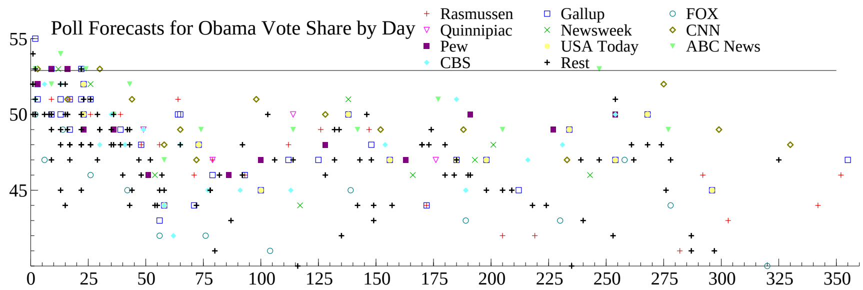

Figure 1: Plot of poll forecasts for Obama vote share in 2008 Presidential Election by day.

companies are represented by different colors and symbols. Gallup is one of the most frequent pollsters and its polls (blue empty squares) appear to become more accurate as election day nears. All polls throughout the campaign appear from Figure 1 to underpredict Obama's eventual vote share, and even in the final few days the majority of polls announced suggest a vote share lower than what eventually results.

Gelman and King (1993) investigate the observed variability in polls despite the fact that election outcomes are particularly predictable at the outset of campaigns. They find that voters learn over the campaign which contributes to some extent to the variability of polls, meaning that early polls are less reliable relative to those conducted nearer to election date. Additionally their research suggests that poll forecasts should be dominated at all stages by expert opinion, statistical and other types of forecast models that embody some subset of that information. Gelman and King focus on Presidential elections, and the 1988 election in particular, although it is undoubtedly the case that many of their conclusions generalize. Nonetheless they do note that some of the effects they emphasize will likely be different for primary elections, and presumably for Senate, Governor and House races.

3.2 Prediction Markets

Prediction markets are markets in which participants buy and sell contracts in event (including election) outcomes. For example, if the market was the 2008 Presidential Election, the contracts would be for the Democratic candidate to win, or the Republican candidate to win. Prediction markets have attracted a great deal of attention in recent years from academic economists because, as Berg et al. (2008) note, their primary role is as a forecasting tool rather than a resource allocation mechanism (although to some extent it can be said that they are part of the portfolio allocation problem of participants since there exists some a priori expected rate of return). Nonetheless, provided their design mechanism is effective, the prices produced will reflect expected probabilities of outcomes. Furthermore, in general the markets are short-term (the majority of those we consider in this paper last for considerably less than a year), and once the outcome is realized, the true value of the contract is known. This property enables researchers to consider whether or not prediction markets forecast events well.

Berg et al. (2008) note the important differences between polls and prediction markets as forecasting devices. The former, at least in principle, are representative samples of the population (or deliberately selected sections of the population), whereas prediction markets are self-selected in that market participants must actively choose to take part. As a result, prediction market participants are anything but representative of the general population; as Berg *et al* point out, “traders are typically young, white, well educated and have high family incomes”. Nonetheless, it is clear that this ought to be irrelevant for the accuracy of prediction market forecasts since the payoff structure means that market participants must put aside their own particular preferences over candidates and predict the voting behaviour of the electorate at large if they are to make non-negative returns.

Prediction Markets (PMs) have been up and running for over 20 years at this point; with the first market having been established for the 1988 Bush-Dukakis contest (Rothschild, 2009). Berg et al. (2008), in their meta-analysis of the performance of PMs in elections in the USA and in other established democracies, found that, in terms of predicting the final result, “in the majority of (...) cases the market does about as well as the average poll, sometimes worse but often better, even if by a small margin” (p. 747); a finding that builds on a previous papers comparing PMs data to poll data. Erikson and Wlezien (2008) note that electoral markets have gained intellectual traction both in academic circles and in the popular press, with Surowiecki’s (2004) *The Wisdom of Crowds* popularizing the idea that aggregated predictions of voting outcomes, which ask individuals to evaluate likely electoral outcomes can be ‘better’ (p. 35) than polls, which ask voters how they themselves will vote. Indeed, futures markets have been extended to predict non-electoral political phenomena, most controversially the likelihood of terrorist attacks (Wolfers and Zitzewitz, 2004).

Dissenting voices have questioned the alleged superiority of election markets to polls, and some of the most recent published research comparing polls to electoral markets has sought to ‘discount’ (Rothschild, 2009) or ‘de-bias’ (Erikson and Wlezien, 2008) poll data, in order to account for observed early poll margin overestimation and anti-incumbency biases in polling data. However, while Erikson and Wlezien found that de-biased poll data outperform national-level electoral market data for US Presidential elections between 1988 and 2004, especially in winner-takes-all predictions, Rothschild found that de-biased market-based data outperforms de-biased poll data in state-level forecasts in the 2008 US Presidential and Senatorial elections. Additionally, Lee and Moretti (2009) use a model of Bayesian learning to suggest that information passes from polls to PMs, while Sjöberg (2009) also challenges the notion of the ‘wisdom of crowds’ by looking at a range of different groups of forecasters for Swedish elections. Finally, similar comparisons of prediction market forecasts to more traditionally generated forecasts have been carried out in sports betting, looking at prices posted by bookmakers against prediction markets (Spann and Skiera, 2009; Croxson and Reade, 2011; Franck et al., 2011, see, for example). As such, the relative performance of poll versus electoral market data is still open to debate.

In the next three sections we introduce in turn the three prediction markets (*Iowa Electronic Markets*, *Intrade* and *Betfair*) we will examine for their performance in predicting election outcomes.

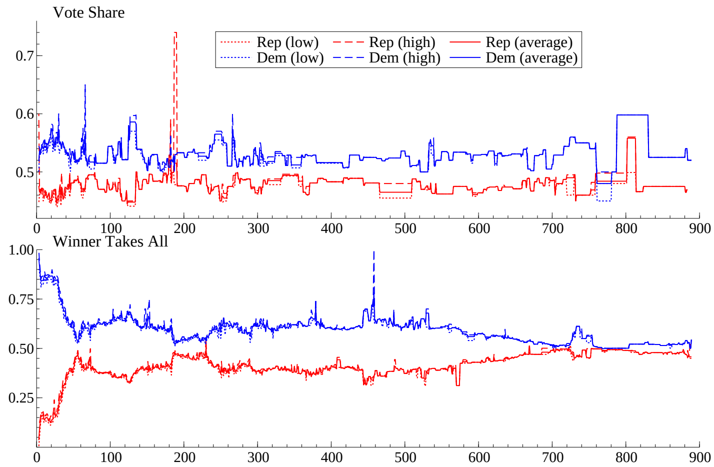

Figure 2: Vote share (top panel) and winner takes all (bottom panel) markets for 2008 US Presidential Election on IEM.

3.2.1 Iowa Electronic Markets

Iowa Electronic Markets (IEM) are not-for-profit operated prediction markets generally linked to political elections (but also markets have existed for box office movies and other one-off events), and have been running since 1988. On IEM participants are limited in their exposure in any trade to \$500. Markets have been set up for all major elections for over a decade, and in particular we have collected data on their prediction markets since 2000. Their markets tend to have two forms, either a winner-takes-all (WTA) or a vote-share (VS) format. The former corresponds to forecasting the election outcome and hence would be described as $\widehat{W}_{i,j,f,T|t}$, while the latter corresponds to providing forecasts of the form $\widehat{V}_{i,j,f,T|t}$.

We have 45,590 observations covering 38 elections; those elections are Presidential (2000, 2004 and 2008), Congressional (House and Senate and a joint market, 2000–2010), and a somewhat ad hoc collection of mayoral elections and primary elections (alongside Democratic and Republican Conventions since 2000). See Table 6 for more details. For each market and contract IEM makes available on a daily basis the number of trades (units and dollar volume), the highest and lowest prices traded at, and the average price. An important distinction between IEM markets for House and Senate elections is that the contracts bought and sold are for macro outcomes: Either the Democrats or the Republicans have a majority in the House or Senate as a result of the election. Similarly, the Republican and Democratic Convention markets allow the trading in contracts about the eventual outcome rather than individual primaries. As Table 6 shows, there are a couple of exceptions (e.g. New York Senate), but generally IEM does not provide markets for individual elections outside Presidential elections.

Figure 2 presents prices from IEM for the 2008 Presidential election; on the top panel the VS market prices are plotted (high, low and average), while on the bottom panel the WTA prices are plotted. These two graphs visualize the difference between the vote share type of forecasts that polls constitute, $\widehat{V}_{i,j,f,T|t}$, and the probability of outcome, $\widehat{W}_{i,j,f,T|t}$, that prediction markets usually provide. Viewing from right-to-left, as the election day draws near, although the vote shares forecast don't diverge particularly strongly (top panel), the probability of each outcome does diverge substantially, and in the final days of the election the probability of a Democratic victory is around 85% and above.

Berg et al. (2008) compare IEM to polls for Presidential elections back to 1988, and find that IEM outperforms the polls in head-to-head comparisons. In relation to their study, we consider a much broader selection of recent elections of all types for both polls and IEM (again see Tables 6 and 5), a strategy that affords us a larger dataset of more recent polls. Berg et al. (2008) note how the demographic of market participants on IEM has changed over the years since 1988, and hence by considering elections only after 2000 we expect to have data more representative of the current IEM demographic.

Erikson and Wlezien (2009) on the other hand contends that polls are actually more informative than prediction markets making use of novel data on informal prediction markets for presidential elections going back to 1880, and using multivariate methods. Although we cannot match Erikson and Wlezien (2009) for sample length, we have substantially broader depth in that we consider here various types of election other than Presidential elections giving us the sample size we mentioned above. Additionally, we have multiple observations per election whereas Erikson and Wlezien (2009) only use one observation, taken immediately prior to each election. Given King and Gelson's findings regarding the accuracy of polls as election date nears, it seems likely that this is a favourable comparison for polls; we will be able to shed light on this using our dataset. Erikson and Wlezien (2008) do consider polls with longer time horizons until the election when considering Presidential elections between 1988 and 2004, and conduct an empirical bias correction for these polls.

3.2.2 Betfair

Betfair is an online betting company providing markets primarily in sports events but also increasingly in political events such as elections. In the jargon, participants either *back* or *lay* bets on events, equivalent to buying or selling contracts paying out contingent on that event happening, such as a politician to win an election. Betfair operates a limit order book, as it matches participants willing to buy and sell contracts at particular prices. In contrast to IEM, participants are not restricted in their potential exposure on Betfair to any arbitrarily imposed limit, and Betfair is a for-profit company; it seems likely that this would influence the self-selection that takes place for potential market participants. When applied to our context of election outcomes with two parties, Betfair yields observations corresponding to $\widehat{W}_{i,j,f,T|t}$.

As our objective is to consider what publicly available information could be used to best forecast an election, although our data is very rich, only certain aspects of it are relevant. Market participants using Betfair can see what prices are available to buy and sell contracts in an event, and how much money (liquidity) is available at each price (buy or sell).

Figure 3 shows the evolution of the implied probabilities (reciprocal of the market prices for contracts) for each party to win the 2008 Presidential election over the 900 days prior to the election. In Figure 3 the maximum and minimum

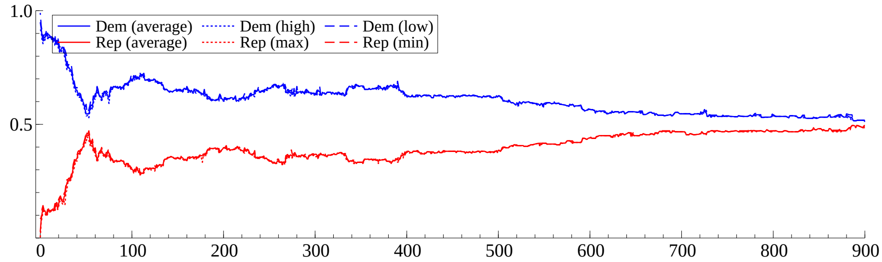

Figure 3: Plot of implied probability of each party winning 2008 Presidential Election from Betfair.

prices for a given day, as well as the average price, are plotted, but as these are very similar to each other it is almost impossible to distinguish them in the plot. This plot can be compared to the bottom plot in Figure 2 which shows the same price evolution for IEM. As with IEM, the Democrats are always the favourites throughout the 900 days shown, and by a slightly larger margin consistently than IEM, with a similar pattern of divergence in probabilities in the final 50 days of the campaign. These plots suggest that, as with polls highlighted by Gelman and King, also with prediction markets learning takes place and the nearer an election is, the more decided becomes the market on the most likely outcome.

3.2.3 Intrade

Intrade was a prediction market specializing in US political elections.[8] There are no limits on the amount that individuals can trade, as opposed to IEM, and the format is essentially identical: market participants trade contracts whose payout is contingent on some event occurring.[9] As such, when thinking specifically about election outcomes with two parties, our Intrade data provides us with observations corresponding to $\widehat{W}_{i,j,f,T|t}$.

We have data from the 2004 and 2008 US Presidential Election; for both years we have all individual state voting and for 2008 we have a range of additional politically related markets.

For the 2004 elections, we have daily data consisting of the high, low and closing prices, while in 2008 we have data on individual trades carried out on the exchange. The 2008 data provides information on whether contracts were bought or sold, the price at which the trade took place and the quantity, alongside a timestamp of when the trade took place. We have 29,196 observations from the 2004 Presidential election (although all of these relate to individual state markets rather than the overall outcome), and 411,858 from the 2008 elec-

[8]Indeed, the perception has long been that Intrade provides for US elections while Betfair does so for UK elections; see, for example, http://www.midasoracle.org/2007/04/24/betfair-vs-tradesports-intrade/ (last accessed April 17 2012).

[9]Servan-Schreiber et al. (2004) compare Intrade to News Futures, a prediction market based on 'play money', using a 'game' format, to ascertain whether "money matters". They find that money doesn't appear to improve the forecast performance of prediction markets. Our analysis, comparing Intrade and Betfair to IEM will shed some light on this question since IEM restricts the amount of money participants are able to bet.

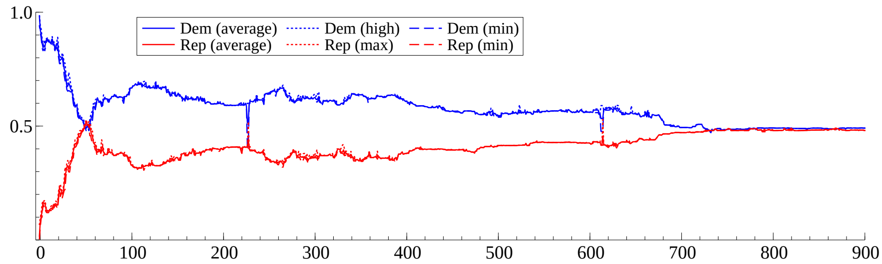

Figure 4: Plot of implied probability of each party winning 2008 Presidential Election from Intrade.

tions (although not all of these relate specifically to elections — see Table 8 for a breakdown).

The purpose of this study is to find the best forecast method, and since Intrade reports on its website very visibly the price of the last agreed trade, clearly more information exists for 2008 for us to assess the Intrade predictions, but nonetheless the information from 2004 does provide additional information.

Figure 4 shows the Intrade implied probabilities (prices divided by 100) for the same 900 days prior to the 2008 Presidential election as in Figures 2 and 3. As is perhaps clear, the two parties appear a little closer as measured by Intrade; at two years prior to the election, the two are absolutely even, and even with just 50 days to go before the election, the two implied probabilities overlap for a short period. It is quite likely that this overlap with just 50 days remaining was due to market manipulation; one trader apparently traded so as to raise the price on Intrade for the Republican candidate, John McCain.[10] As discussed by Hanson and Oprea (2009), we do not see this as necessarily a problem for our analysis; a manipulator in a liquid market might be viewed as offering other traders a kind of free lunch in correcting that manipulator's attempts to distort.

As with both previous prediction markets, Intrade has also attracted academic interest; Gil and Levitt (2007) investigated market efficiency looking at the 2002 FIFA World Cup, while Hartzmark and Solomon (2012) considered the disposition effect using NFL markets. Snowberg et al. (2007) use TradeSports, incorporated by Intrade in 2008, to infer implications from elections onto the macroeconomy by using the 2004 US Presidential election when unreliable exit polls caused substantial price variation within a single day.

We can now compare all three markets on one plot, in Figure 5, and over a long period of time the co-movement between these series is very clear. A comparison between these three prediction markets is of great interest, not least because the self-selection that takes place in each market will likely be different; Betfair does not allow those based in the US to trade in their markets, while IEM does operate in the US but restricts its participants in their exposure, while Intrade allows those based in the US to trade in its markets but does not restrict the exposure of participants. Hence it is of interest to compare these

[10] See http://marginalrevolution.com/marginalrevolution/2008/10/manipulation-of.html for more information on this.

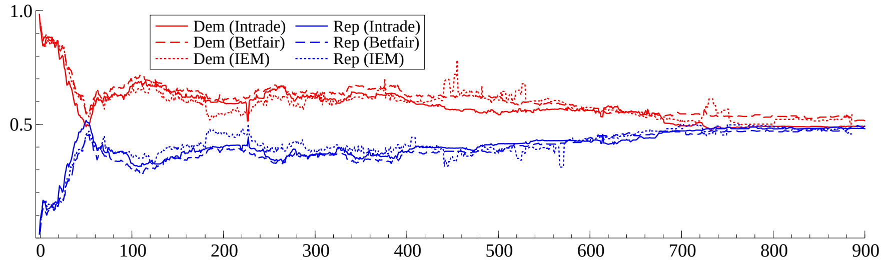

Figure 5: Plot of comparable prices from all three prediction markets from the 2008 Presidential Election.

three markets in their ability to forecast elections; do these differences matter? Furthermore can any of them, as Berg et al. (2008) assert, improve upon polls?

3.3 Statistical Models

Statistical models have been employed on a number of occasions to predict US election outcomes; Fair (1996), for example, provides a readily updated dataset from which to predict outcomes outcomes of presidential and congressional elections. His model relies on a small number of predictors based on economic (real growth and inflation), past political (results, incumbent identity, terms served) and geopolitical outcomes (wars), and is estimated using ordinary least squares.

We create *real-time* forecasts and compare them to our other forecast models. This step is important because macroeconomic data is subject to often sizeable revisions, meaning that the data that would have been used to construct election forecasts in previous years may have been subsequently revised. Making use of the archival economic data resource *Alfred* from the Federal Reserve of St Louis (http://alfred.stlouisfed.org/), we construct datasets for each election since 1996 as would have been available at the time. This is clearly important for any comparison between forecasts different methods as our opinion polls and prediction market data are all based on real-time information.

In Figure 6 we present the forecasts that our real-time replication of Ray Fair's statistical model generate for elections since 1996 at both the presidential and congressional level. The forecast variable is the two-party vote share for the Republican party. When we compare forecasts from our statistical model to the other forecast models, rather than running ordinary least squares estimation on vote share outcomes, we run a linear probability model on a binary outcome variable which is 1 if the Republican candidate wins in an election, 0 otherwise. This way our forecast is a probabilistic one rather than a vote share forecast.

3.4 Expert Opinion

A large amount of information regarding any particular election is difficult to quantify; into such a vacuum, expert opinion can play an important role. By expert opinion we refer to forecasts expressed by interested commentators decreed to be experts. Cast in terms of the traditional forecasting literature we

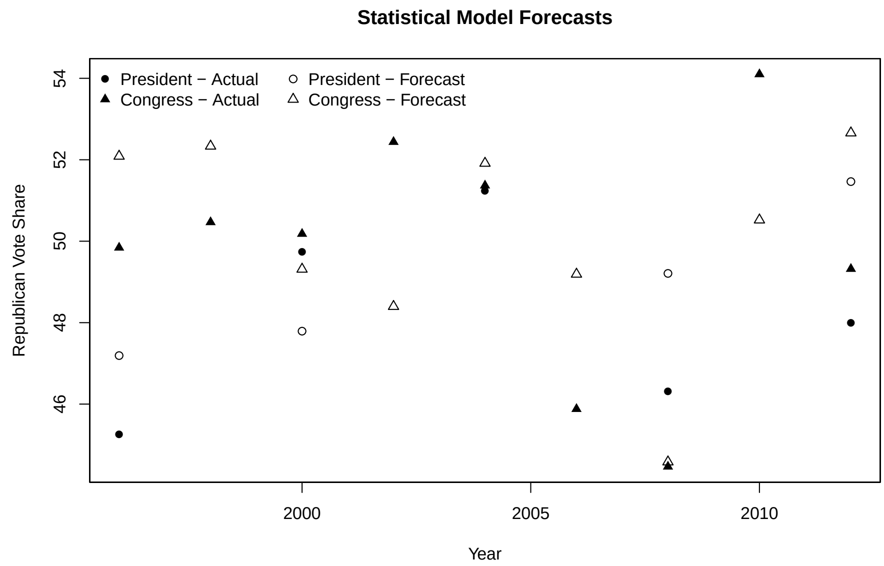

Figure 6: Plot of forecasts from our statistical model, and outcomes from Presidential and Congressional elections since 1996.

might refer to these kinds of forecasts as judgemental forecasts (for a review, see Lawrence et al., 2006).

We consider a dataset of expert forecasts collected by Graefe et al. (2012), freely available via the PollyVote project.[11] This dataset covers the 2004, 2008 and 2012 presidential elections, and is composed of up to sixteen anonymised forecasters who are allowed to make regular updates to their forecasts.

Figure 7 shows expert forecasts for the 2008 presidential election, and specifically for the Republican two-party vote share. Each symbol represents a different individual forecast at a different time horizon in advance of the election, and the dotted line is the actual outcome. In grey we also plot the average of these expert forecasts; at least graphically it appears better than many of the individual forecasts; Graefe et al. (2012) argue that the average represents a better forecast than any individual expert chosen at random.

As with opinion polls and our statistical model, the expert forecasts are in terms of vote shares and hence need converting to probabilities of outcomes. We use a simpler variant of the method employed for opinion polling, using only the forecast and number of days as explanatory variables in a linear probability model.

4 Methodology

We seek metrics to assess each candidate forecast model. Such metrics should be impartial between the different forecast models and hence give us an objective outcome regarding the best forecast model. An immediate obstacle in this

[11]See http://pollyvote.ifkw.uni-muenchen.de/en/publications/

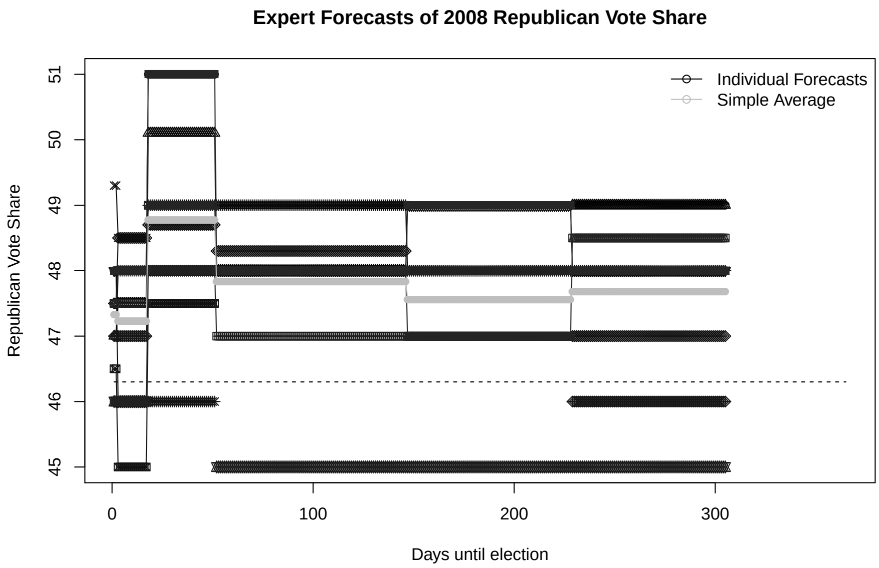

Figure 7: Plot of forecasts from sixteen experts for the Republican vote share at the 2008 Presidential Election.

pursuit is that we have two types of forecasts: vote shares, $\widehat{V}_{i,j,f,T|t}$, and probabilistic forecasts, $\widehat{W}_{i,j,f,T|t}$. With the former the outcome is continuous over the unit interval, whereas for the latter the outcome is a binary variable. Various methods exist to turn opinion poll vote shares into outcome probabilities; Page (2008) develops a theoretical parametric approach, while alternatives include non-parametric density estimation using historical poll outcomes. In recent elections, Nate Silver has been converting polling information into probabilistic forecasts, although as would be expected with a commercial endeavour, details regarding the methods employed are undisclosed.[12]

We employ a regression method to create probabilities; regression methods as simple as ordinary least squares (OLS) allow the estimation of the probability associated with observing binary outcomes such as election outcomes (O_{ij}). Alternatively we could use a probit or a logit model to transform our polling shares. This is the method employed by Vaughan Williams and Reade (2014) to transform opinion polls into outcome probabilities.

A probit regression model using O_{ij} as dependent variable estimates, using historical data, the probability of election victory conditional on a number of explanatory factors, which we denote by the vector X_{ij}. We write the model as:

$$\mathsf{P}\left(V_{ij} = \max_k V_{ik} \middle| X_{ij} \right) = \Phi\left(\beta X_{ij}\right), \tag{2}$$

where β is a vector of coefficients, and Φ is the cumulative density function of the standard normal distribution. In X_{ij} we include information from each

[12]See, for example, http://goo.gl/gv7xQ1, last accessed 16 October 2014.

poll for candidate j in election i. We include candidate j's polled vote share in election i, but we also include the opinion poll vote shares of other candidates in the election, as well as information the poll, such as the sample size used. It seems sensible to explore whether such additional dimensions of polls provide systematic information on the likelihood of victory of a candidate polled in addition to the polled vote share; not least doing so enables us to correct for any known biases in opinion polling relative to known outcomes. Such bias correction is in line with the method proposed by Erikson and Wlezien (2008) of correcting poll outcomes before comparing with prediction markets.

Using the estimated coefficients from (2) we then construct predicted probabilities for election outcomes given opinion poll shares; this is a forecast exercise, forecasting the likelihood of a candidate winning given information in one particular opinion poll, $\mathsf{P}\left(V_{ij} = \widehat{\max_k V_{ik}} \,|X_{ij}\right)$. We estimate the model only using data up to the most recent day of polling, and then forecast the probability for forthcoming polls. We do this so that the transformation from polling share to probability only relies upon information known at the time of the poll.[13]

As with other attempts to compare forecast methods such as Erikson and Wlezien (2009), if we rely on direct comparison forecast by each forecaster for particular events, we will be severely restricted in our number of elections and hence observations relative to the total datasets we have at our disposal. Instead we assess each forecasting method over all the elections we are able to collect data on *for that method* (see Tables 5–8). Thus we attempt to establish for each forecaster, independent of the others, how well it forecasts election outcomes, before comparing these performances between forecasters. There is considerable overlap between our datasets for each forecaster such that we are considering forecast performance over very similar datasets.[14]

Any forecast assessment is reliant on the loss function assumed; what loss do we suffer if the forecast is wrong in a particular direction? With elections and vote shares, such a loss function is unlikely to be symmetric since if a forecast is for 51%, then if the outcome is ±2%, the direction is fundamental — up and the election outcome (in terms of vote share) is unaltered, down and the outcome changes.

Our objective is to pick the winner in a forthcoming election, and hence a very simple metric for forecasts is accuracy: how often does the forecast outcome occur? Hence whether the forecast is for a vote share of 51% or 65% is somewhat irrelevant provided that that event happens. However, it is likely that election outcomes that are nearer to 50% (for a two-party election) will induce lower success rates. Bearing this in mind, and given that often US Presidential elections are very close, we also move to consider forecasts more generally. In this sense, it must be the case that a good forecast is both unbiased, displaying no systematic biases, and precise, and we will outline how we test for this in Sections 4.1 and 4.2.

[13] We make use of an additional dataset of opinion polls for the US presidential elections in 2000 and 2004 to estimate our baseline conversion model.

[14] While two or three of our models will have overlapping observations for many elections, the elections for which we have comparable data for all four models is restricted to essentially the 2008 Presidential election (for example, for Intrade we have only state-level 2004 Presidential election data but for Betfair and IEM we have only national level market data).

4.1 Accuracy and Unbiasedness

A very simple measure of accuracy is the percentage of correct forecasts. This is the most direct measure of what minimises our loss function probabilistically; the forecast method that forecasts correctly most often must on average yield the lowest loss. In the case of markets that provide $\widehat{W}_{i,j,f,T|t}$ forecasts, we take a forecast to be predicting a particular outcome if that particular forecast probability is the highest of the candidates in an election. Hence we take, for $i = 1, \ldots, N$ contestants in an election $\widehat{W}^*_{i,j,f,T|t} = \max_k \widehat{\mathsf{P}}_t (V_{k,j,T} > 0.5)$, the candidate or party with the highest forecast probability of winning, as the forecast outcome at that point. For forecasts of the nature $\widehat{V}_{i,j,f,T|t}$, we take $\widehat{V}^*_{i,j,f,T|t} = \max_k \widehat{V}_{k,j,T|t}$, the maximum vote share, as the favourite and hence predicted outcome. We also denote $V^*_{i,j,T|t} = \max_k V_{k,j,T|t}$ as the candidate with the highest vote share and hence the winner of the popular vote in an election.[15]

We thus calculate, for forecast model f, the percentage of correct forecasts as:

$$\%_f = \frac{\sum_{i=1}^{N_i} \sum_{j=1}^{N_j} \sum_{t=1}^{N_t} 1_{\left\{V^*_{i,j,T} = \widehat{V}^*_{i,j,f,T|t}\right\}}}{N_i N_j N_t}, \tag{3}$$

where N_j is the number of elections considered, N_i the number of candidates and N_t the number of time periods. We compare forecasts along this dimension to assess the *accuracy* of polls.[16]

A related but distinctly different concept to accuracy is that of *unbiasedness*. An *unbiased* forecast can be defined separately for each kind of forecast:

- An unbiased vote share forecast is, on average, equal to the true vote share outcome: $\mathsf{E}\left(\widehat{V}_{i,j,f,T|t}\right) = V_{i,j,T}$.
- An unbiased probability forecast is, on average, equal to the true probability that that candidate wins the election: $\mathsf{E}\left(\widehat{W}_{i,j,f,T|t}\right) = W_{i,j,T}$.

Hence forecasts that are accurate *can also be biased*, provided the bias is in the correct direction; if polls are consistently upward biased for candidates that eventually win, then despite being biased they will be very accurate in predicting the outcome, whereas polls that are consistently downward biased for candidates that eventually win will be very inaccurate as well as biased.

When we consider vote share forecasts for candidate i in election j, $\widehat{V}_{i,j,f,T|t}$, after an election has happened we observe the true $V_{i,j,T}$ and hence we can evaluate the *forecast error*:

$$\widehat{e}^V_{i,j,f,T|t} = V_{i,j,T} - \widehat{V}_{i,j,f,T|t}. \tag{4}$$

We can use this forecast error to consider the possibility of biased forecasts. In taking the simple average of (4) we thus learn whether or not a forecast method

[15] Which, as noted earlier, need not correspond to the actual winner of the election.

[16] We could be more demanding with our measure of accuracy in (3) and require that forecast models got the final ranking of candidates correct; all we require is that the forecast model correctly identifies the favourite. Given that the majority of US elections are two-party, this distinction is unlikely to be of much practical relevance.

is unbiased or not. Hence we calculate:

$$MFE_t = \sum_{i=0}^{N_{f,t}} \hat{e}^V_{i,j,f,T|t}\,, \tag{5}$$

where $N_{f,t}$ denotes the number of forecasts we have for each forecast method at each time period t.

Although we could substitute W for V in (4) and (5) when we observe forecasts that are probabilities of outcomes, it is likely that because the outcome is binary that a summation such as in (5) would unfairly penalize probabilistic forecasts that are above 50% but not by particularly much. As a result, we convert vote share forecasts into probabilistic forecasts.

In addition, we use a Mincer-Zarnowitz test, or a *calibration test*, to further evaluate unbiasedness. That is, we regress the outcome on the probability as produced by the forecast method:[17]

$$W_{i,j,T} = \alpha_W + \beta_W \widehat{W}_{i,j,f,T|t} + \varepsilon^W_{i,j,t}. \tag{6}$$

The assumption we place on $\varepsilon_{i,j,T}$ determines the kind of regression model we employ; although it can be shown that estimating (6) via OLS induces heteroskedasticity, it is most convenient for our analysis to estimate using OLS since that implies an iid (independent and identical distribution) assumption for $\varepsilon^W_{i,j,t}$ of $\varepsilon^W_{i,j,t} \sim (0, \sigma^2_W)$; we will later make use of this.

In (6), if $\alpha_V = 0$ and $\beta_V = 1$ the forecast method is said to be *unbiased* since then $\mathsf{E}\left(\widehat{W}_{i,j,f,T|t}\right) = W_{i,j,T}$ (because $\mathsf{E}\left(\varepsilon^W_{i,j,t}\right) = 0$). Hence the F-test of the null hypothesis $\alpha_V = 0$ and $\beta_V = 1$ is our test of *unbiasedness*. If the constant term $\alpha_V \neq 0$ then the forecast method does exhibit some systematic bias, while if $\beta_V \neq 1$ then if the true probability of an event changes, the forecast method either over- or under-adjusts. This phenomenon is commonly referred to in the betting literature as the *favourite longshot bias* (FLB). The conventional FLB ($\beta_W > 1$) is where bettors relatively over-bet event outcomes with lower implied probabilities of winning (inferred from the odds) and relatively (though not necessarily absolutely) under-bet event outcomes with higher implied probabilities of winning. The reverse FLB ($\beta_W < 1$) occurs where bettors relatively over-bet and under-bet the converse.

We seek a method to assess forecasts that is unitless due to the two different types of forecast in our dataset, and hence we employ the same method outlined in (6) when considering forecasts from polls. Just as $\alpha_V = 0$ and $\beta_V = 1$ implies $\mathsf{E}\left(\widehat{W}_{i,j,f,T|t}\right) = W_{i,j,T}$, and hence that the fitted line through the scatter plot of forecasts against outcomes corresponds to the 45 degree line, we can apply the same methodology to polls; does a polled vote share of, say, 47% imply that on average the resulting outcome is 47%? Hence we run the regression of:

$$V_{i,j,T} = \alpha_V + \beta_V \widehat{V}_{i,j,f,T|t} + \varepsilon^V_{i,j,t}. \tag{7}$$

Equivalently to above, $\alpha_V = 0$ and $\beta_V = 1$ imply that on average polled levels equal actual outcomes and hence the forecast model is unbiased: $\mathsf{E}\left(\widehat{V}_{i,j,f,T|t}\right) =$

[17] Note we write $W_{i,j,T}$ here, adding a t to the outcome; the observed outcome does not change through time, we just add this in order that we can run regressions for different forecasts at different time ts before the election occurs.

$V_{i,j,T}$. The value of this method in comparing our two types of forecast is that for probabilistic forecasts ($\widehat{W}_{i,j,f,T|t}$) we compare to the expected value of, $\mathsf{E}(W_{i,j,T})$ rather than the binary variable itself, $W_{i,j,T}$. This reduces a potential distortion when comparing forecast errors from vote shares and probablistic forecasts.

Additionally, if $\alpha_V = 0$ and $\beta_V = 1$ are imposed then $\widehat{e}^V_{i,j,f,T|t} = \widehat{\varepsilon}^V_{i,j,f,T|t}$, our regression model (7) becomes equivalent to the forecast error from (4) earlier for vote shares, and hence we can think about (7) as a generalized forecast error. By running the regression in (7) we learn about the actual relationship between $\mathsf{E}\left(\widehat{V}_{i,j,f,T|t}\right)$ and $V_{i,j,T}$ rather than asserting that the two are equal. Similarly as with (6), if $\beta_V > 1$ we have FLB: the favourite on average gets a higher vote share than the outsider.

Thus in both regression models, (6) and (7), the null hypothesis of $\alpha_g = 0$ and $\beta_g = 1$, $g \in \{V, Y\}$, implies that the forecast method is unbiased — on average it forecasts without error. Although a visual examination of the estimated α and β coefficients will be informative, it is also useful to construct a *direct test of unbiasedness*, and hence we use an F test of the hypothesis that $\alpha_g = 0$ and $\beta_g = 1$ to evaluate the unbiasedness of our forecast methods. Because in both types of forecast the F-test measures departures from unbiasedness (expected values), it should not be influenced by the distinction between $V_{i,j,T}$ being continuous on the unit interval and $W_{i,j,T}$ being binary.

4.2 Precision

Having considered unbiasedness, it is now helpful to move on to thinking about precision — how precise are the forecasts we get? A conventional measure of the precision of a forecast is the *mean squared forecast error* (MSFE) — squaring the forecast errors we calculated in (4) and summing:

$$MSFE_g = \sum_{i=0}^{N_{f,t}} \left(\hat{\varepsilon}^g_{T|t}\right)^2, \quad g \in \{V, Y\}, \tag{8}$$

This is an approximation to the variance of the forecast, centred around the outcome, and hence is equivalent to the estimated standard error for our regression model, (7), denoted $\widehat{\sigma}^2_f$ since the formula for that is:

$$\widehat{\sigma}^2_g = \frac{1}{N_{g,t}} \sum_{i=1}^{N_{g,t}} (\widehat{\varepsilon}^g_{i,j,t})^2 = \frac{1}{N_{f,t}} \sum_{i=1}^{N_{g,t}} \left(V_{i,j,T} - \widehat{\alpha}_g - \widehat{\beta}_g \widehat{V}_{i,j,f,T|t}\right)^2 = \sum_{i=0}^{N_g} \left(\hat{e}^g_{T|t}\right)^2 = MSFE_g, \quad g \in \{V, Y\}, \tag{9}$$

provided $\widehat{\alpha}_Y = 0$ and $\widehat{\beta}_Y = 1$. Thus $\widehat{\sigma}^2_g$ is a more general measure of forecast accuracy than MSFE which imposes restrictions on (7).

In essence, $\widehat{\sigma}^2_W$ measures how imprecise a prediction market is at providing probabilistic forecasts, while $\widehat{\sigma}^2_V$ measures how imprecise at providing vote-share forecasts a poll is. However, these two $\widehat{\sigma}^2$ measures consider the precision around the *actual* relationship between forecasts and outcomes, rather than the 45-degree line (which $\alpha_g = 0$ and $\beta_g = 1$ would imply). The equivalent MSFE measures impose the $\alpha_g = 0$ and $\beta_g = 1$ restrictions without testing

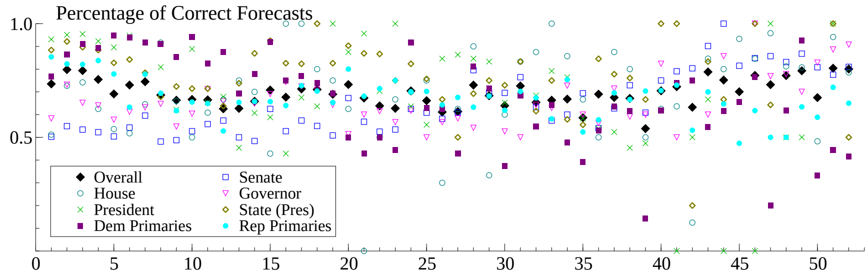

Figure 8: Percentage of polls that correctly predict election outcome by weeks until election.

their appropriateness but nonetheless do provide important information — how dispersed around the 45-degree line are the forecasts.

Hence we assess forecast errors and the F test of $\alpha = 0$ and $\beta = 1$ to assess *forecast unbiasedness* and analyze MSFEs and $\widehat{\sigma}^2$ assess *forecast precision*.

5 Results

We now consider the accuracy, bias and precision of each of our forecasting models. We first consider *accuracy* via the percentage of correct forecasts (3), presenting the results graphically for each market then assessing the markets head-to-head, before considering via regression methods the accuracy and precision of the markets.

5.1 Accuracy

Figure 8 reports the overall percentage of opinion polls that correctly forecast the actual outcome by weeks before the election was due to take place, and the bars represent the number of polls that fall into each category.[18] We then refine by particular types of election.

The overall percentage of forecasts which are correct drawn from polls is 71.0%, increasing to 76.8% if only the Presidential elections of 2004 and 2008 are considered. In Figure 8 we chart the performance of polls as the distance to election, and hence the forecast horizon, increases. There appears to be no particular improvement in poll performance as an election nears, something which contrasts with Gelman and King's findings. Even for presidential elections, for which performance does appear to peak in the 4–5 weeks before an election, performance is actually comparable if not better between weeks 19 and 23 where just under 100 polls record a success percentage of slightly over 90%. The black diamonds in Figure 8 show the overall performance of polls for all elections we consider, and this does improve slightly from a low of just above 60% with 13 weeks remaining to around 80% with two weeks remaining, but this

[18]Where we point out that the 'winner' in terms of vote share is taken to be the candidate that won the most votes, hence for example in the 2000 Presidential election, Gore is classed as the winner as he won more of the popular vote.

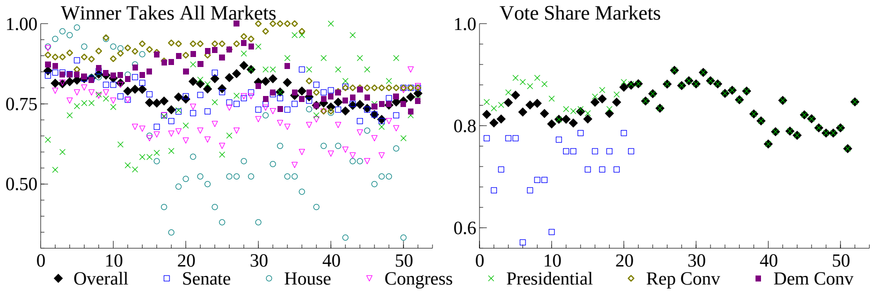

Figure 9: Percentage of IEM prices that correctly predict election outcome by weeks until election.

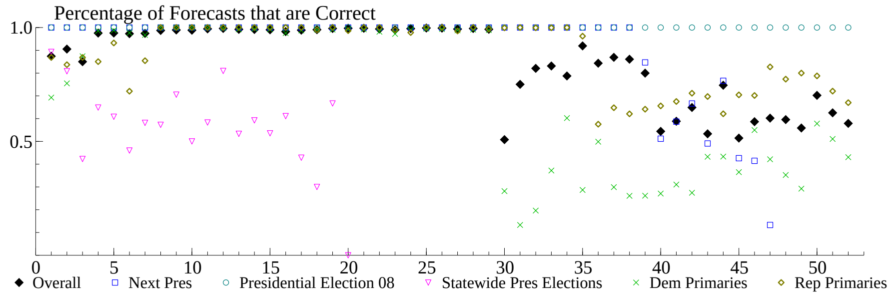

Figure 10: Percentage of Betfair prices that correctly predict election outcome by weeks until election.

performance is not significantly better than polling performances 30–40 weeks before an election.[19]

Turning to IEM, we split forecasts into vote share (VS) and winner-takes-all (WTA) markets. The percentage of forecasts that were correct is 83.5% (VS) and 79.8% (WTA), changing to 85.2% and 72.3% respectively for presidential elections. Figure 9 presents a graphical breakdown of how these percentages move as the time until election increases with WTA on the left panel, VS on the right. What is perhaps most notable here is that IEM's WTA Presidential elections forecasts seem significantly worse than all of its other election forecasts in the final 20 weeks before the election takes place — before that, its success ratio is considerably higher the IEM average. Intriguingly, the VS markets show Presidential election forecasts better than average, although there is little other than Presidential markets in this category of market (85% of our observations are from the 2000, 2004 and 2008 Presidential elections), but in absolute terms the percentage is higher, remaining just shy of 85% up to the final week of the election, as opposed to the WTA percentage of between 55% and 65%.

With Betfair, we display the percentage of correct forecasts in Figure 10 weekly for a year in advance of each election. Betfair's overall percentage of correct forecasts is 85.5% and 99.8% for the 2008 Presidential election (falling to 95.3% for the Next President market), while we present the breakdown by

[19]In Figure 13 we plot standard error bands helping to make such a comparison.

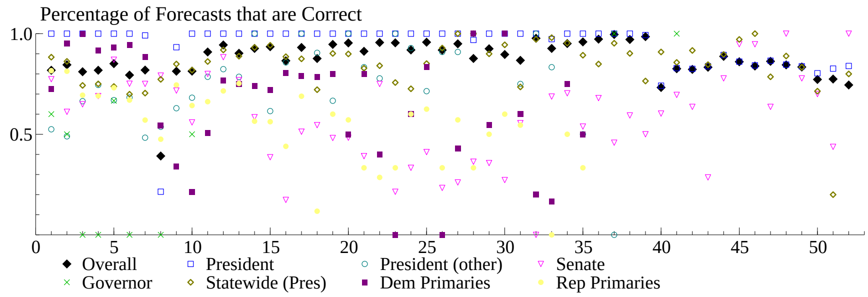

Figure 11: Percentage of Intrade prices that correctly predict election outcome by weeks until election.

weeks before an election occurs in Figure 10. Aside from this almost perfect record in forecasting the 2008 election, additionally for both Republican and Democratic primaries, Betfair has a success percentage of 87.1% in the final 30 weeks of campaigns. Betfair forecasts of statewide elections for the electoral collage (upside down triangles) improve from essentially zero 20 weeks before election day to 90% in the final week.

Figure 11 shows Intrade percentages in the same format as for the previous three candidate forecast models. Intrade has a percentage of forecasts turning out correct of 84.0%, rising to 88.1% for the 2008 Presidential election. As with Betfair, we see a high level of correct forecasts, particularly for the 2008 Presidential election where again up to around week 38, Intrade prices imply a correct forecast almost every trade. From Table 8 we have a large collection of markets from Intrade related to US Presidential elections in 2008 other than simply the outcome or vote share, and the purple dots in Figure 11 represent these; as can be seen, the prediction record on these more eclectic events (e.g. whether a particular video will be released by the LA Times by a particular date) is dramatically worse than for US elections, as even in the week before the election takes place the percentage of forecasts that are correct is only around 50%. However as these minor markets make up a small fraction of our total observations, their impact on the overall percentage is minimal; it is lower percentages for statewide Presidential elections, Governor markets and other Presidential-related markets that pulls the overall percentage down.

Considering expert opinion, we take a forecast probability of victory of above 50% as indicating the event an expert is predicting to happen. We find that 93% of the time experts correctly predicted the outcome of a Presidential election. Spread over the weeks before an election, this varies little; Figure 12 shows that it is actually at longer horizons that experts perform best; between 21 and 38 weeks before an election, experts predict perfectly the outcome of elections. It is not clear that this is consistent with an enlightened voter theory regarding election outcomes since the experts become less accurate in forecasting the nearer an election is; to be consistent with the theory, experts should continue to be as accurate as the election nears while voters become more learned.

Finally considering our statistical model approach, we only have 14 forecasts, and of those 8 are correct (in that both the forecast and outcome are above (below) 50), giving an accuracy rate of 57%.

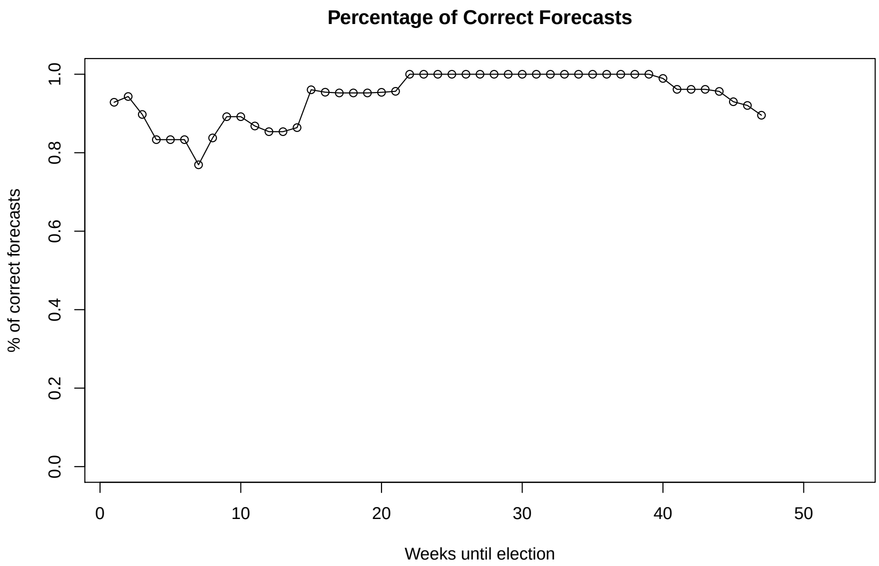

Figure 12: Percentage of expert predictions that correctly predict election outcome by weeks until election.

It is informative to compare our poll and prediction market candidate forecasts with each other directly on a plot, and Figure 13 does that, plotting the four as different series over the year up to elections. We additionally include 95% significance-level standard error bounds around each market's plot, enabling us to assess whether differences are significant.[20] The plot indicates that the best forecasting method to choose based on accuracy (percentage of forecasts turning out correct) is Betfair from 30 weeks before an election up to election week. It is at this kind of range that the expert forecasts (plotted separately) are also essentially always correct. In the three weeks immediately before an election the performance of the four methods becomes much less dispersed, but nonetheless Betfair remains significantly better than polls and Intrade, although not significantly better than IEM WTA. With the exception of three weeks (8, 30 and 40 weeks prior), polls are dominated by prediction markets in providing accurate forecasts in the 40 weeks before an election occurs. In the final 10 weeks before an election, the performance of the IEM (both VS and WTA) and Intrade markets is indistinguishable statistically, and with the exception of forecasts 2 and 3 weeks before an election, significantly superior to polls.

Thus, concluding our discussion of accuracy in terms of the percentage of correct forecasts, we find that prediction markets and expert opinion dominate polls in providing accurate forecasts.

[20]The distinctly differently sized confidence bands is more a function of sample size rather than any inherent uncertainty in particular models. This is because we only have one observation per day per market for IEM, only a relatively small number of polls per market per week, whereas we have often hundreds and even thousands of trades per day on Intrade and Betfair. We do not reduce our Intrade or Betfair samples down to any kind of daily average in order not to discard any important data.

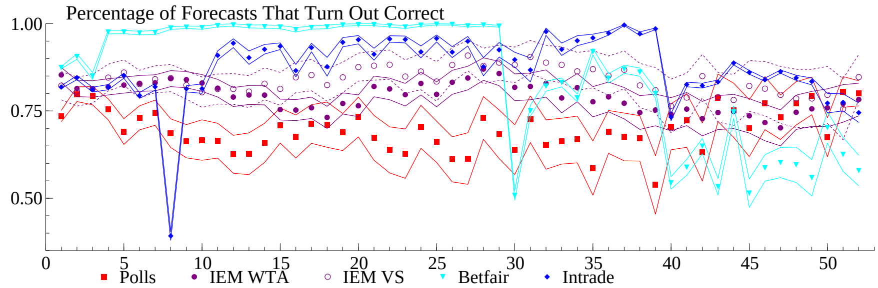

Figure 13: Plot comparing the percentage of correct forecasts for poll and prediction market forecasts of forecast information for US elections.

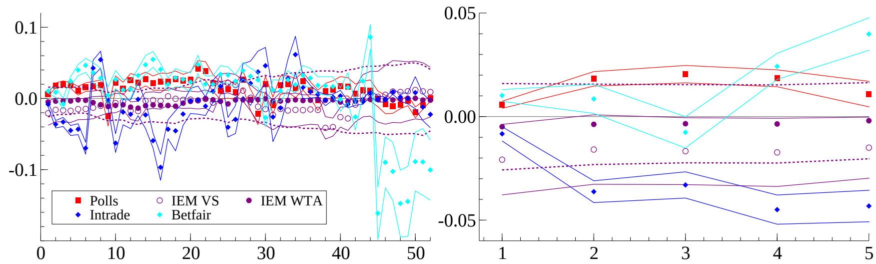

Figure 14: Plot comparing average forecast errors, as calculated in (5), head-to-head between opinion polls and prediction markets, by weeks until election. Left panel is all weeks in the year prior to an election, right panel focuses on final five weeks pre-election.

5.2 Bias and Precision

We next consider *bias*, whether the expected value of a forecast equals the true value, and *precision*, how much variance a forecast model exhibits, graphically before conducting a basic regression analysis help quantify our findings.

Figure 14 plots the average errors with standard error bounds for opinion polls and prediction markets hence giving an idea about the *bias* of forecasts. It is worth noting that the standard error bounds contain information on the *precision* of each forecast since the standard error of the forecast error is equal to the squared root of the MSFE (from (9)) with the restriction $\alpha_g = \beta_g - 1 = 0$ imposed. The left plot shows the entire year before an election, while the right plot zooms in on the final five weeks.

From the left panel, over short intervals all forecast models display biases in one direction or another, but over the longer term these biases do appear to cancel each other out. The polls and IEM (both) deliver what appears to be the most consistent performance, with Betfair and Intrade fluctuating markedly around zero. In general polls have a slight upward bias, while IEM has a slight downward bias, moreso in VS than WTA, Betfair upward and Intrade downward. It thus appears that prediction markets provide higher forecast success yet are not necessarily less biased than polls.

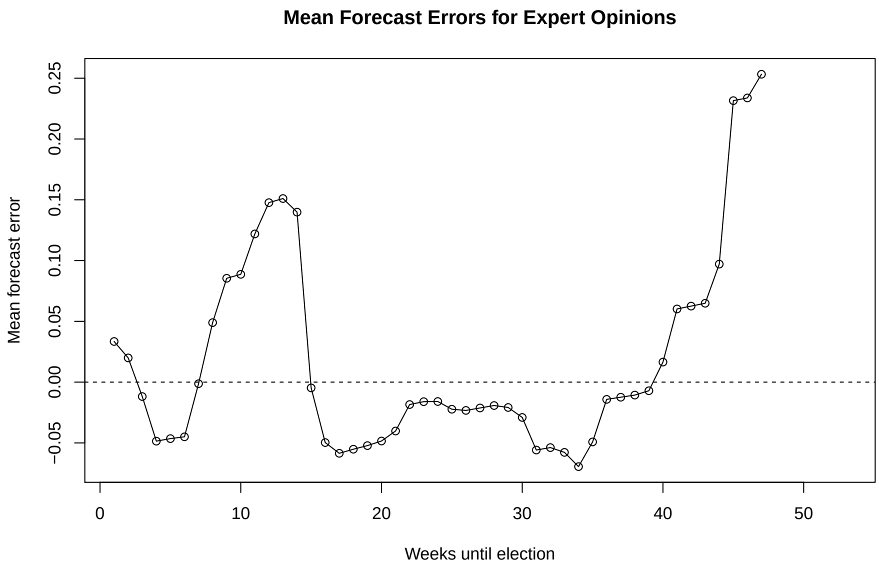

Figure 15: Plot of average forecast errors, as calculated in (5), for expert opinion.

We plot expert opinion in Figure 15, for comparison with opinion polls and prediction markets. Apart from around 10 weeks before an election, expert opinion appears competitive with polls and prediction markets, with errors fluctuating around zero but never more than about ± 0.05. Similar to Betfair, expert opinion appears to be much worse further than 40 weeks before the election. Considering finally our statistical model, the average forecast error across all forecasts is 0.055, which is competitive with our other forecasts. The statistical forecasts are all conducted in real-time terms one week before the election, as this is when the most recent data release was to that particular election.

Considering the regression model approach outlined in Section 4, Table 1 contains the output from the regression models for all of our models over all available observations.

The regression in columns (1) is of our opinion polls transformed into outcome probabilities, while column (2) differs slightly in that it is a regression of vote share forecasts ($V_{i,j,T}$) from IEM on vote share forecasts ($\widehat{V}_{i,j,f,T|t}$), while the remaining columns contain regressions of actual outcomes of elections ($W_{i,j,T}$) on implied probabilities ($\widehat{W}_{i,j,f,T|t}$) from prediction markets, expert opinion and our statistical model. The principle is the same in all regressions; unbiased forecasts should be reflected in finding that $\alpha_g = 0$ and $\beta_g = 1$, namely that the implied regression line is on the 45 degree line and hence a poll forecasting a vote share on average is correct (columns (1) and (2)), and a contract priced implying a particular probability pays out with that frequency (columns (3)–(5)).

The first row of numbers in each column contains the estimates for α_g, the intercept coefficient, while the second row contains the estimates for β_g, the slope coefficient. Beneath these coefficients is the output of an F-test of $\alpha_g = 0$ and $\beta_g = 1$; the first line is the p-value, the probability of a incorrect rejection of

	(1) Polls	(2) IEM (V)	(3) IEM	(4) Intrade	(5) Betfair	(6) Expert	(7) Stat.
$\widehat{\alpha}_g$	-0.003***	0.135***	0.082***	-0.020***	-0.062***	-0.095***	0.133
	(-0.557)	(62.113)	(31.031)	(-28.363)	(-47.351)	(-15.040)	(0.363)
$\widehat{\beta}_g$	0.999***	0.622***	0.842***	1.044***	1.189***	1.179***	0.849
	(77.082)	(56.241)	(105.674)	(516.245)	(460.119)	(133.37)	(1.284)
p_val	0.427	0.000	0.000	0.000	0.000	0.000	0.896
F_stat	0.8507	1943.101	482.209	409.130	2849.682	225.5	0.111
$\widehat{\sigma}_g^2$	0.432	0.042	0.139	0.102	0.113	0.247	0.501
T	21803	11429	31737	356620	183775	9086	14

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 1: Regressions for bias and precision for all forecast models over all observations.

the null hypothesis, and the second row is the F-test statistic itself. In essence, the larger is the F-test statistic, the further away from $\alpha = 0$ and $\beta = 1$ is that particular set of forecasts. Because of the huge sample sizes of our regressions (from the final row), it is expected that p-values will be very small, with the exception of the final column for our statistical model.[21] The largest F-test statistics by some distance are for IEM vote share in the second column, and Betfair in the fifth column. For IEM this is mainly driven by a departure from unity of the $\widehat{\beta}_g$ coefficient, at 0.622, and the constant coefficient at 0.135, while for Betfair it would appear more a function of sample size since the deviation from $\alpha_g = \beta_g - 1 = 0$ is smaller yet the sample size is almost ten times as large as for polls. In terms of actual coefficient sizes, the smallest departure from $\alpha_g = \beta_g - 1 = 0$ is for opinion polls and Intrade. Opinion polls in particular do not reject the null hypothesis of $\alpha_g = \beta_g - 1 = 0$; that said, the opinion polls in this regression have been bias corrected as part of the transformation to probabilistic forecasts. However, this bias correction was done on a real-time basis such that the appropriate transformation could have been made at the time and used as a forecast; to this extent we feel the comparison is a fair one.[22]

In terms of precision, the vote share regression of IEM has a much smaller $\widehat{\sigma}_g$ than the winner-takes-all regressions; within the winner takes all Betfair and Intrade are more precise than IEM, with Intrade appearing most precise. Expert opinion appears less precise than prediction markets, but transformed polls are even less precise than experts; the statistical model provides the least precise forecasts.[23]

Finally we note that IEM (both VS and WTA) and the statistical model display evidence of a reverse favourite-longshot bias (FLB), whereas Betfair,

[21](Campos et al., 2003) discuss this problem with inference in large samples, and suggest adjusting significance level to $T^{-0.8}$, where T is sample size.

[22]In a previous version of the paper we considered this regression with simple polling vote shares, and found some evidence of the kinds of biases that a strategy such as ours was able to correct for. It is worth noting also that similarly one could consider bias correcting prediction market prices.

[23]Again, in a previous version of the paper the regression here for poll shares provided a very low $\widehat{\sigma}_g$, suggesting that the size of this parameter is a function of the kind of forecast being made.

	(1) Polls	(2) IEM (V)	(3) IEM	(4) Intrade	(5) Betfair	(6) Expert	(7) Stat.
$\widehat{\alpha}_g$	0.103***	0.105***	0.067***	-0.009***	-0.059***	-0.141***	0.133
	(6.922)	(4.926)	(5.530)	(-3.418)	(-23.634)	(-3.505)	(0.363)
$\widehat{\beta}_g$	0.647***	0.855***	0.868***	1.003***	1.130***	1.266***	0.849
	(19.834)	(8.413)	(28.222)	(200.593)	(290.260)	(22.573)	(1.284)
p_val	0.000	0.000	0.000	0.000	0.000	0.000	0.896
F_stat	68.977	12.618	16.328	12.015	584.645	13.325	0.111
$\widehat{\sigma}_g^2$	0.453	0.128	0.111	0.127	0.082	0.280	0.501
T	3467	360	1110	43295	41882	293	14

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2: Regressions for bias and precision for all relevant forecasts for observations within one week of an election.

Intrade and expert opinion exhibit FLB. The traditional favourite-longshot bias is the observed phenomenon, found in numerous studies dispersed in time and across the world, that 'longshots' (outcomes quoted at high odds) tend to win less often than implied in the odds while 'favourites' (outcomes quoted at low odds) tend to win relatively more often than implied in the odds (e.g. Sung and Johnson (e.g. 2010); Snowberg and Wolfers (e.g. 2010). This may help explain the difference in accuracy noted in Figure 13, where Betfair and Intrade appear to dominate the other three forecast models. If favourites win more often than their forecast suggests, as with Betfair and Intrade, then the percentage of forecasts that turn out correct must be higher, and vice versa with polls and IEM.

Thus our regression models from Table 1 lend support to the conclusions drawn from Figure 14: over the longer horizon all forecast models appear to exhibit quite substantial bias, with the exception of Intrade and opinion polls, while opinion polls appear to provide the least precise forecasts.

As with *accuracy*, we can refine somewhat our analysis of *bias* and *precision* by looking at forecasts made at various points before an election. Tables 2–4 show the same regressions as Table 1 but for forecasts made within the final week of an election (Table 2), forecasts made between 2 and 10 weeks before and election (Table 3), and forecasts made between 11 and 40 weeks before an election (Table 2).

From Table 2 polls display much more bias in the final week than in the overall regressions in Table 1, while Intrade continues to display little bias. IEM markets still display reverse FLB while Betfair still displays a small FLB. The performance of expert opinion seems little different in the final week compared to overall (and to the other forecast models), while the statistical model, being only estimated over real-time data from the final week before the election, is unchanged from Table 1.

As we move to longer horizons in Tables 3 and 4, we observe that both Intrade and Betfair display a more pronounced bias, IEM markets show no clear pattern, polls less bias and expert opinion bias patterns are essentially unchanged from the final week. Intriguingly the regressions coefficients for expert opinion and Intrade are almost identical at the 2–10 week horizon, although

	(1) Polls	(2) IEM (V)	(3) IEM	(4) Intrade	(5) Betfair	(6) Expert
$\widehat{\alpha}_g$	-0.082***	0.124***	0.051***	-0.082***	-0.179***	-0.087***
	(-9.708)	(17.194)	(11.412)	(-32.560)	(-63.753)	(-4.515)
$\widehat{\beta}_g$	1.177***	0.784***	0.943***	1.136***	1.422***	1.136***
	(64.086)	(21.271)	(74.193)	(218.537)	(282.062)	(44.182)
p_val	0.000	0.000	0.000	0.000	0.000	0.000
F_stat	50.072	149.530	70.016	533.289	3602.357	14.740
$\widehat{\sigma}_g^2$	0.409	0.136	0.121	0.150	0.076	0.341
T	9626	3218	9597	82135	35054	2499

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3: Regressions for bias and precision for all four firms over observations within 2 and 10 weeks of an election.

both are distinctly different from the other two prediction markets (Betfair and IEM). This suggests a similar pattern in terms of bias, but Intrade displays a much higher degree of precision.

The information presented here suggests that firstly in terms of correct forecasts, prediction markets and expert opinion dominate polls. Consideration of bias and precision shows that all forecast models are shown to be biased in different directions and magnitudes at different times (with the exception of bias-corrected polls over longer horizons than the final week before an election), while levels of precision also vary with prediction markets being the most precise and polls and expert opinion the least.

Returning to the work of Gelman and King (1993), we note that certainly in terms of accuracy expert opinion (and prediction markets) do dominate opinion polls, which is a finding of theirs, and particularly at longer horizons. Furthermore, our finding that in the final week of campaigns even bias-corrected polls perform notably worse without a similar fall off in expert (or prediction market) forecasts is also consistent, as it suggests that the most learning on the part of voters is taking place during that period.

Considering also Erikson and Wlezien (2008), like them we bias correct polls, which they argue is essential in order to appropriately compare polls with prediction markets. Nonetheless, despite bias correcting we find that opinion polls in the final week before an election display marked bias and a lack of precision relative to prediction markets, and overall perform worse in terms of accuracy.

6 Conclusions

In this paper we have investigated a number of information sources that might be used to form a forecast of an election outcome. We consider the forecasts of opinion polls, three commonly used prediction markets, a statistical model, and expert opinion. We assess these forecast models in terms of accuracy, bias and precision. We make use of very large datasets recording the forecast performance of these different models over a large number of elections in the US.

	(1) Polls	(2) IEM (V)	(3) IEM	(4) Intrade	(5) Betfair	(6) Expert
$\widehat{\alpha}_g$	0.027***	0.124***	0.087***	-0.069***	-0.088***	-0.117***
	(2.643)	(23.438)	(24.190)	(-46.585)	(-44.505)	(-17.66)
$\widehat{\beta}_g$	0.969***	0.708***	0.831***	1.289***	1.301***	1.201***
	(42.053)	(26.797)	(73.972)	(281.890)	(296.885)	(127.47)
p_val	0.005	0.000	0.000	0.000	0.000	0.000
F_stat	5.308	274.713	292.717	2007.052	2509.161	26.932
$\widehat{\sigma}_g^2$	0.442	0.140	0.141	0.079	0.113	0.243
T	7584	6485	17317	67949	81133	5359

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: Regressions for bias and precision for all four firms over observations within 11 and 40 weeks of an election.

Our analysis suggests that prediction markets tend to provide more accurate forecasts, and although corrected poll forecasts appear less biased they are less precise. In particular commercial prediction markets display distinct favourite longshot bias, suggesting that they are more able to identify favourites that subsequently win the election, which helps explain why these models forecast more accurately. Expert opinion similarly displays bias and is not particularly precise.

Overall our results provide more evidence on the relative performance of opinion polls and prediction markets; support is given to Gelman and King (1993) as opinion poll performance is dominated, particularly in the final week before an election, and while bias corrected poll do perform competitively at a general level, we nonetheless fail to find convincingly in favour of opinion polls in this assessment, which contrasts with the findings of Erikson and Wlezien (2009).

References

J. Bates and C. W. J. Granger. The Combination of Forecasts. *Operations Research Quarterly*, 20:451–468, 1969.

J.E. Berg, F.D. Nelson, and T.A. Rietz. Prediction market accuracy in the long run. *International Journal of Forecasting*, 24(2):285–300, 2008.

J. Campos, D.F. Hendry, and H.-M. Krolzig. Consistent Model Selection by an Automatic Gets Approach. *Oxford Bulletin of Economics and Statistics*, 65 (s1):803–819, December 2003.

K. Croxson and J.J. Reade. Exchange vs. Dealers: A High-Frequency Analysis of In-Play Betting Prices. Discussion Papers 11-19, Department of Economics, University of Birmingham, January 2011.

R.S. Erikson and C. Wlezien. Are Political Markets Really Superior to Polls

As Election Predictors? *Public Opinion Quarterly*, 72(2):190–215, Summer 2008.

R.S. Erikson and C. Wlezien. Markets vs. Polls as Predictors: An Historical Assessment of US Presidential Elections. In *Annual Meeting of the American Association for Public Opinion Research, Hollywood, Florida*, 2009.

Ray C Fair. Econometrics and presidential elections. *The Journal of Economic Perspectives*, pages 89–102, 1996.

E. Franck, E. Verbeek, and S. Nüesch. Sentimental Preferences and the Organizational Regime of Betting Markets. *Southern Economic Journal*, 78(2): 502–518, 2011.

A. Gelman and G. King. Why are American Presidential Election Campaign Polls so Variable when Votes are so Predictable? *British Journal of Political Science*, 23(4):409–451, October 1993.

R. Gil and S. Levitt. Testing the Efficiency of Markets in the 2002 World Cup. *The Journal of Prediction Markets*, 1:255–270, 2007.

A. Graefe, J.S. Armstrong, R.J. Jones, and A.G. Cuzan. Combining forecasts: An application to elections. Annual meeting paper, APSA, February 2012.

R. Hanson and R. Oprea. A Manipulator can Aid Prediction Market Accuracy. *Economica*, 76(302):304–314, 2009.

Samuel M Hartzmark and David H Solomon. Efficiency and the disposition effect in NFL prediction markets. *The Quarterly Journal of Finance*, 2(03), 2012.

F.A. Hayek. The Use of Knowledge in Society. *The American Economic Review*, 35(4):510–530, September 1945.

W. Hurley and L. McDonough. A Note on the Hayek Hypothesis and the Favorite-Longshot Bias in Parimutuel Betting. *The American Economic Review*, 85(4):949–955, September 1995.

S. Kou and M.E. Sobel. Forecasting the Vote: A Theoretical Comparison of Election Markets and Public Opinion Polls. *Political Analysis*, 12(3):277–295, 2004.

M. Lawrence, P. Goodwin, M. O'Connor, and D. Önkal. Judgmental forecasting: A review of progress over the last 25 years. *International Journal of Forecasting*, 22(3):493–518, 2006. ISSN 0169-2070. doi: http://dx.doi.org/10.1016/j.ijforecast.2006.03.007.

D.S. Lee and E. Moretti. Bayesian Learning and the Pricing of New Information: Evidence from Prediction Markets. *The American Economic Review*, 99(2): 330–336, 2009.

A. Leigh and J. Wolfers. Competing Approaches to Forecasting Elections,: Economic Models, Opinion Polling and Prediction Markets. *Economic Record*, 82(258):325–340, September 2006.

L. Page. Comparing prediction market prices and opinion polls in political elections. *Journal of Prediction Markets*, 2(1):91–97, May 2008.

P.W. Rhode and K.S. Strumpf. Historical Presidential Betting Markets. *Journal of Economic Perspectives*, 18(2):127–142, Spring 2004.

D. Rothschild. Forecasting Elections. *Public Opinion Quarterly*, 73(5):895–916, 2009.

E. Servan-Schreiber, J. Wolfers, D.M. Pennock, and B. Galebach. Prediction Markets: Does Money Matter? *Electronic Markets*, 14(3):243–251, 2004.

Lennart Sjöberg. Are All Crowds Equally Wise? A Comparison of Political Election Forecasts by Experts and the Public. *Journal of Forecasting*, 28(1): 1–18, 2009. ISSN 1099-131X. doi: 10.1002/for.1083. URL `http://dx.doi.org/10.1002/for.1083`.

V.L. Smith. Markets as Economizers of Information: Experimental Examination of the "Hayek Hypothesis". *Economic Inquiry*, 20(2):165–179, April 1982.

E. Snowberg and J. Wolfers. Explaining the Favorite-Longshot Bias: Is It Risk-Love or Misperceptions? *Journal of Political Economy*, 118(4):723–746, 2010.

E. Snowberg, J. Wolfers, and E. Zitzewitz. Partisan impacts on the economy: evidence from prediction markets and close elections. *The Quarterly Journal of Economics*, 122(2):807, 2007.

Wolfers J. Snowberg, E. and E. Zitzewitz. Information (In)Efficiency in Prediction Markets. In L. Vaughan Williams, editor, *Information Efficiency in Financial and Betting Markets*, pages 366–386. Cambridge University Press, Cambridge, 2005.

M. Spann and B. Skiera. Sports forecasting: A comparison of the forecast accuracy of prediction markets, betting odds and tipsters. *Journal of Forecasting*, 28(1):55–72, 2009. ISSN 1099-131X. doi: 10.1002/for.1091. URL `http://dx.doi.org/10.1002/for.1091`.

M. Sung and J.E.V. Johnson. Revealing Weak-Form Efficiency in a Market for State Contingent Claims: The Importance of Market Ecology, Modelling Procedures and Investment Strategies. *Economica*, 77:128–147, 2010.

J. Surowiecki. *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations*. Little, Brown, 2004. ISBN 0-316-86173-1.

L. Vaughan Williams and J.J. Reade. US Elections: Evaluating the Recent Performance of Prediction Markets and Opinion Polls. Technical report, Department of Economics, University of Reading, 2014. mimeo.

J. Wolfers and E. Zitzewitz. Prediction Markets. *Journal of Economic Perspectives*, 18(2):107–126, 2004.

A Data Information Tables

The Tables on the following pages contain information on the make-up of each of our datasets introduced in Section 3 and analysed in Section 5.

Democratic Primaries	Governor 2006	Governor 2010	House 2006	House 2010	House 2010 (cont)	House 2010 (cont)	Republican Primaries 2008	Senate 2006	Senate 2010	State 2008
Alabama	Alabama	Alabama	Arizona 8	Alabama 2	Indiana 8	New York 2	Arizona	Arizona	Alabama	Alabama
Arizona	Alaska	Alaska	Colorado 4	Alabama 5	Indiana 9	New York 4	California	Connecticut	Alaska	Alaska
California	Arizona	Arizona	Colorado 7	Arizona 1	Iowa 1	New York 13	Colorado	Maryland	Arizona	Arizona
Connecticut	Arkansas	Arkansas	Connecticut 2	Arizona 3	Iowa 2	New York 19	Connecticut	Michigan	Arkansas	Arkansas
Delaware	California	California	Connecticut 4	Arizona 5	Iowa 3	New York 20	Delaware	Minnesota	California	California
Democratic	Colorado	Colorado	Connecticut 5	Arizona 7	Kansas 3	New York 22	Florida	Missouri	Colorado	Colorado
Florida	Connecticut	Connecticut	Florida 13	Arizona 8	Kentucky 3	New York 23	Georgia	Montana	Connecticut	Connecticut
Georgia	Florida	Florida	Florida 16	Arkansas 1	Kentucky 6	New York 24	Illinois	New Jersey	Delaware	Delaware
Illinois	Georgia	Georgia	Florida 22	Arkansas 2	Louisiana 2	New York 25	Iowa	Ohio	Florida	Florida
Indiana	Idaho	Hawaii	Georgia 8	Arkansas 4	Louisiana 3	New York 27	Massachusetts	Pennsylvania	Georgia	Georgia
Iowa	Illinois	Idaho	Georgia 12	California 3	Maine 1	New York 29	Michigan	Rhode Island	Hawaii	Hawaii
Kansas	Iowa	Illinois	Illinois 6	California 11	Maine 2	North Carolina 2	Minnesota	Tennessee	Idaho	Idaho
Kentucky	Kansas	Iowa	Illinois 8	California 18	Maryland 1	North Carolina 4	Missouri	Virginia	Illinois	Illinois
Maryland	Maine	Kansas	Indiana 2	California 20	Massachusetts 4	North Carolina 7	Nevada	Washington	Indiana	Indiana
Massachusetts	Maryland	Maine	Indiana 8	California 45	Massachusetts 5	North Carolina 8	New Hampshire	Senate Elections 2008	Iowa	Iowa
Michigan	Massachusetts	Maryland	Indiana 9	California 47	Massachusetts 6	North Carolina 11	New Jersey	Alaska	Kansas	Kansas
Minnesota	Michigan	Massachusetts	Iowa 1	Colorado 3	Massachusetts 10	North Carolina 13	New York	Colorado	Kentucky	Kentucky
Missouri	Minnesota	Michigan	Kentucky 3	Colorado 4	Michigan 1	Ohio 1	Republican Nomination	Georgia	Louisiana	Louisiana
Nevada	Nebraska	Minnesota	Kentucky 4	Colorado 7	Michigan 7	Ohio 6	South Carolina	Kentucky	Maryland	Maine
New Hampshire	Nevada	Nebraska	Minnesota 6	Connecticut 1	Michigan 9	Ohio 10	Tennessee	Louisiana	Missouri	Maryland
New Jersey	New Hampshire	Nevada	Nevada 2	Connecticut 2	Michigan 15	Ohio 12	**Republican Primaries 2012**	Maine	Nevada	Massachusetts
New Hampshire	New Mexico	New Hampshire	Nevada 3	Connecticut 4	Minnesota 1	Ohio 13	Alabama	Minnesota	New York (Dio Vs Gil)	Michigan
New York	New York	New Mexico	New Hampshire 2	Connecticut 5	Minnesota 6	Ohio 15	Arizona	Mississippi Special	New York (Tow Vs Sch)	Minnesota
North Carolina	Ohio	New York	New Jersey 7	Florida 2	Minnesota 7	Ohio 16	Colorado	Nebraska	North Carolina	Mississippi
Ohio	Oklahoma	Ohio	New Mexico 1	Florida 8	Minnesota 8	Ohio 18	Florida	New Hampshire	North Dakota	Missouri
Oregon	Oregon	Oklahoma	New York 20	Florida 12	Mississippi 1	Oklahoma 2	Georgia	New Jersey	Ohio	Montana
Pennsylvania	Pennsylvania	Oregon	New York 24	Florida 22	Mississippi 2	Oregon 1	Hawaii	New Mexico	Oklahoma	Nebraska
South Carolina	Rhode Island	Pennsylvania	New York 26	Florida 24	Mississippi 4	Oregon 4	Illinois	North Carolina	Oregon	Nevada
Tennessee	South Carolina	Rhode Island	New York 29	Florida 25	Missouri 3	Oregon 5	Iowa	Oregon	Pennsylvania	New Hampshire
Texas	Tennessee	South Carolina	North Carolina 11	Georgia 2	Missouri 4	Pennsylvania 3	Louisiana	Virginia	Utah	New Jersey
Virginia	Texas	South Dakota	Ohio 1	Georgia 8	Nebraska 2	Pennsylvania 4	Maine		Vermont	New Mexico
West Virginia	Vermont	Tennessee	Ohio 15	Georgia 12	Nevada 3	Pennsylvania 6	Massachusetts		Washington	New York
Wisconsin	Wisconsin	Texas	Ohio 18	Hawaii 1	New Hampshire 1	Pennsylvania 7	Michigan		West Virginia	North Carolina
	Wyoming	Utah	Pennsylvania 6	Idaho 1	New Hampshire 2	Pennsylvania 8	Minnesota		Wisconsin	North Dakota
		Vermont	Pennsylvania 7	Illinois 8	New Jersey 3	Pennsylvania 10	Mississippi			Ohio
		Wisconsin	Pennsylvania 8	Illinois 10	New Jersey 6	Pennsylvania 11	Nevada			Oklahoma
		Wyoming	Pennsylvania 10	Illinois 11	New Jersey 12	Pennsylvania 12	New Hampshire			Oregon
			Texas 22	Illinois 14	New Mexico 1	Pennsylvania 13	Ohio			Pennsylvania
			Washington 8	Illinois 17	New Mexico 2	Pennsylvania 15	Oklahoma			Rhode Island
			Wisconsin 8	Indiana 2	New Mexico 3	Pennsylvania 17	South Carolina			South Carolina
					New York 1	Rhode Island 1	Tennessee			South Dakota
						South Carolina 5	Vermont			Tennessee
						Tennessee 4	Virginia			Texas
							Washington			Utah
										Vermont
										Virginia
										Washington
										West Virginia
										Wisconsin
										Wyoming

Table 5: Summary of data publicly available from *Real Clear Politics* on polling for US elections.

Election	Start	Finish	Freq.	Percent
Congress 2000	28jan1999	08nov2000	2,592	5.69
Congress 2002	19jul2002	07nov2002	448	0.98
Congress 2004	17jun2004	05nov2004	568	1.25
Congress 2006	01jun2006	12nov2006	664	1.46
Congress 2008	22aug2008	07nov2008	312	0.68
Congress 2010	24nov2009	04nov2010	1,344	2.95
Senate Elections 2004	17jun2004	05nov2004	426	0.93
Senate Elections 2006	01jun2006	10nov2006	492	1.08
Senate Elections 2008	22aug2008	07nov2008	234	0.51
Senate Elections 2010	24nov2009	04nov2010	1,003	2.20
Florida Senate Election 2010 (vote share)	04jun2010	30nov2010	720	1.58
Florida Senate Election (winners takes all)	04jun2010	30nov2010	720	1.58
Minnesota Senate Election 2008 (vote share)	20aug2008	08nov2008	243	0.53
Minnesota Senate Election 2008 (winners takes all)	20aug2008	08nov2008	243	0.53
New York Senate Election 2000	14jun1999	08nov2000	2,725	5.98
House Elections 04	17jun2004	05nov2004	426	0.93
House Elections 06	01jun2006	10nov2006	492	1.08
House Elections 08	22aug2008	07nov2008	234	0.51
House Elections 10	24nov2009	04nov2010	1,004	2.20
Presidential Election 2000 (vote share)	03jan2000	05nov2000	920	2.02
Presidential Election 2000 (winners takes all)	24apr2000	10nov2000	597	1.31
Presidential Election 2008 (vote share)	01jun2006	07nov2008	1,830	4.01
Presidential Election 2008 (winners takes all)	01jun2006	07nov2008	1,830	4.01
Presidential Election 2004 (vote share)	20feb2003	31jul2004	7,026	15.41
Presidential Election 2004 (winners takes all)	26may2004	05nov2004	426	0.93
Democratic Convention 2000	14jun1999	17aug2000	3,219	7.06
Democratic Convention 2004	20feb2003	30jul2004	3,419	7.50
Democratic Convention 2008	24feb2007	28aug2008	2,300	5.04
Republican Convention 2000	14jun1999	03aug2000	2,339	5.13
Republican Convention 2008	24feb2007	10sep2008	3,141	6.89
Reform Convention 2000	03jan2000	12aug2000	1,105	2.42
Iowa Republican Caucus 12	29aug2011	05jan2012	889	1.95
New York City Mayoral Election 2001	03oct2001	09nov2001	108	0.24
Philadelphia Mayoral Election 2007 (vote share)	02apr2007	02jul2007	534	1.17
Philadelphia Mayoral Election 2007 (winners takes all)	02apr2007	02jul2007	529	1.16
Mexican Presidential Election 2000 (vote share)	01may2000	02jul2000	244	0.54
Mexican Presidential Election 2000 (winner takes all)	01may2000	02jul2000	244	0.54
Total			45,590	

Table 6: Summary of data publicly available from *Iowa Electronic Markets* on polling for US elections.

Type	Market	Observations	Type	Market	O
Presidential	2004	44,462		Republican Candidate	
Election	2008	11,831		California Primary	
	Democratic Candidate	39,012		Florida Primary	
	Alabama Primary	123		Iowa Caucus	
	Alaska Caucus	2	Republican	Michigan Primary	
	Arizona Primary	132	Candidacy	Nevada Caucus	
	Arkansas Primary	4		New Hampshire Primary	
	California Primary	1,564		New Jersey Primary	
	Colorado Caucus	7		South Carolina Primary	
	Connecticut Primary	78		New York Primary	
	Delaware Primary	13		Total	
	Georgia Primary	63		Additional Runners	
	Idaho Caucus	4		Joe Biden	
	Illinois Primary	60		Michael Bloomberg	
	Indiana Primary	1,860		Mike Huckabee	
	Iowa Caucus	941		Mitt Romney	
	Kansas Caucus	6	Next	Ron Paul	
	Kentucky Primary	88	President	Rudy Giuliani	
	Massachusetts Primary	227		Sarah Palin	
	Minnesota Caucus	13		Al Gore	
Democratic	Missouri Primary	184		Barack Obama	
Candidacy	Nevada Caucus	516		Hillary Clinton	
	New Hampshire Primary	1,710		John Edwards	
	New Jersey Primary	250		John McCain	
	New Mexico Caucus	19		Total	
	New York Primary	249		Arkansas	
	North Carolina Primary	551		Indiana	
	North Dakota Caucus	2		New Mexico	
	Ohio Primary	1,041		North Dakota	
	Oklahoma Primary	19		Nevada	
	Oregon Primary	148		Colorado	
	Pennsylvania Primary	1,265	Elections	Florida	
	Tennessee Primary	55	2008	Georgia	
	Texas Primary	2,305		Kentucky	
	Utah Primary	17		Missouri	
	Washington Caucus	69		Montana	
	West Virginia Primary	110		Nebraska	
	Wisconsin Primary	460		Ohio	
	Total	53,167		North Carolina	
				Pennsylania	
				Total	
Grand Total	228,264				

Table 7: Data from Betfair on US Elections

Market	Obs.	Market	Obs.	Market
Presidential Election - Main		*Presidential Election - Other*		*House of Representatives*
Winner (Indiv.)	334,286	Bob Barr - Elec. Coll. Votes	14	2008 House Control
Winner (Party)	18,007	Bob Barr - Popular Vote	444	Dem. Seats in House
Rep. Elec. College Votes	1,225	Dropouts, April	172	Dist. 12 Penn
Electoral College Tie	90	Dropouts, Dec.	210	Dist. 6 Minn
Alabama	38	Dropouts, Feb	519	*Total*
Alaska	209	Dropouts, Jan	349	*Senate*
Arizona	784	Dropouts, Jun	72	2008 Senate Control
Arkansas	248	Dropouts, Jul	1,291	Dem. Seats in Senate
California	293	Dropouts, May	186	Alabama
Colorado	926	Dropouts, Mar	568	Alaska
Connecticut	109	LA Times Obama PLO video	3	Colorado
Delaware	29	Ralph Nader - Popular Vote	117	Georgia
Florida	2,535	Joe Biden to be withdrawn	548	Idaho
New Jersey	405	Sarah Palin to be withdrawn	2,599	Kansas
Nevada	877	Michael Bloomberg Independent	1,073	Kentucky
Nebraska	63	Ron Paul Independent	249	Louisiana
Montana	1,270	Who benefit most from 1st debate	330	Maine
Missouri	2,803	Who benefit most from VP debate	606	Massachusetts
Mississippi	111	Who will run for President?	941	Minnesota
Michigan	487	Date of 1st Debate	215	Nebraska
Minnesota	576	Election Postponed?	48	New Hampshire
Maryland	52	Obama Touch Mkt	9	New Jersey
Georgia	1,605	McCain Touch Mkt	39	New Mexico
Hawaii	17	X: Obama Options. F	123	Mississippi (Class I)
Idaho	16	X: Obama Options. M	15	Mississippi (Class II)
Illinois	57	X: Obama Options. T	1	North Carolina
Indiana	3,956	X: Obama Options. W	83	Oklahoma
Iowa	558	X: Obama Options. W	147	Oregon
Kansas	142	X: Obama Options. W	54	South Carolina
Kentucky	93	X: Obama Options. W	75	South Dakota
Louisiana	194	X: Obama Options. W	98	Texas
Maine	152	X: Obama Options. W	82	Virginia
Massachusetts	26	X: McCain Options. Fr	132	West Virginia
North Carolina	3,212	X: McCain Options. Mo	1	Wyoming (Class I)
North Dakota	1,023	X: McCain Options. We	21	*Total*
Ohio	1,846	X: McCain Options. We	17	*Governor Elections*
Oklahoma	37	X: McCain Options. We	7	Delaware
Oregon	247	X: McCain Options. We	19	Kentucky
Pennsylvania	1,404	X: McCain Options. We	12	Louisiana
New Hampshire	796	X: McCain Options. We	34	Missouri
New Mexico	586	*Total*	11523	North Carolina
New York	118	*Other*		Utah
Rhode Island	30	Immigration Reform Act 20	24	Vermont
South Dakota	120	London Mayoral Election 2008	94	Washington
South Carolina	187	Massachusetts Question 1	8	*Total*
Tennessee	141	Media Endorsements	20	*Party Convention*
Texas	125	New York City Mayoral Term Limits	18	Brokered Conventions
Utah	24	Next Prime Minister of New Zealand	6	Clinton Lifeline
Vermont	61	Next UK Chancellor	15	Hillary Clinton on Dem.
Virginia	2,237	Pres. Job Appr. Rating. Dec 3	119	MI/FL hold new Primar
Washington	128	Pres. Job Appr. Rating. Jun 3	38	Most Superdelegates?
West Virginia	863	Pres. Job Appr. Rating. Mar 3	29	*Total*
Wisconsin	395	Pres. Job Appr. Rating. Sep 3	72	
Wyoming	22	Fairness Doctrine	6	
Total	385,841	*Total*	449	
Total	411,858			

Table 8: Summary of data from Intrade for 2008 elections.